Supplementary Material (ESI) for *Molecular BioSystems*

Electronic Supplementary Information

for

Peptide Mass Fingerprinting Using Isotopically Encoded Photo-crosslinking Amino Acids

Bryan J. Wilkins, Kelly A. Daggett, and T. Ashton Cropp*

Experimental Section

Synthesis of *p*-benzoylphenylalanine[D_{11}]: Reagents and solvents were commercially available and used without further purification. Toluene-d_8 and benzoyl chloride-d_5 were purchased from Sigma. The precursor bromomethylbenzophenone **1**, was synthesized as previously reported[1] and used to assemble the amino acid as follows (Scheme 1). Diethylacetamidomalonate (614 mg, 2.83 mmol) was added to a stirring solution of sodium ethoxide (385 mg, 5.66 mmol) in ethanol (10 mL). **1** (810 mg, 2.83 mmol) was dissolved in ethanol (15 mL) and added dropwise over the course of ten minutes. The reaction mixture was brought to reflux for 24 hours. Upon cooling, water (20 mL) and ether (20 mL) were added and the organic layer was separated. This was washed successively with 1% (w/v) $NaHCO_3$, brine, and then dried over $MgSO_4$ and concentrated under vacuum. This compound was deprotected directly by adding 6 M HCl (40 mL) and then brought to reflux for 24 hours. Upon cooling, the reaction mixture was dried under vacuum. Recrystallization with water afforded the desired amino acid as light yellow crystals (0.501 g, 67%). ^{13}C (MeOD, 100 MHz) σ198.12, 171.03, 140.66, 138.51, 138.05, 133.44, 131.26, 130.59, 130.35, 129.07, 54.72, 36.46. HRMS (ESI) for $C_{16}D_{11}O_3NH_4$ calculated 280.1913, found 280.1928.

GST expression and cross-linking: The gene encoding GST was amplified from pGEX-T4-1 (GE Healthcare) using primers (FWD) 5'- CTAGGATCCCCTATACTAGGTTATTGG-3' and (REV) 5'- CCAGTCGACGCCTCTAGAAACCAGATCCGATTT-3'. The product was then digested with *Bam*HI and *Sal*I and cloned into the *Bgl*II and *Sal*I sites of pBAD-mycHisA (Invitrogen) resulting in appendage of an N-terminal MDPSSR leader peptide and a C-terminal hexa-histidine tag. The mutation of the F51 codon to TAG was performed using standard Quickchange PCR (Stratagene) with primers (FWD) 5'-TTGGGTTTGGAGTAGCCCAATCTTCCT-3' and (REV) 5'-AGGAAGATTGGGCTACTCCAAACCCAA-3'. These mutated plasmids were then

transformed along with the pSUPpBpa plasmid (chloramphenicol marker) into *E. coli* Genehogs (Invitrogen). Transformants containing pBAD/GST with the Phe51 codon mutated to TAG, and pSUPpBpa were inoculated in LB medium (50 mL) containing ampicillin (100 μg/μL) and chloramphenicol (35 μg/μL) and grown in the presence of 1mM pBpa (Bachem), or 1 mM pBpa[D_{11}], or 0.5 mM pBpa/0.5 mM pBpa[D_{11}] and grown to OD_{600}= 0.8. The cultures were induced with 0.2% (w/v) arabinose and allowed to express for 5 hr. Purification of GST was performed using Probond Purification resin (Invitrogen) according to the manufacturer's protocol for native isolation using binding buffer containing imidazole (50 mM NaH_2PO_4, pH 8.0, 10 mM imidazole, 0.5 M NaCl). Purified protein (100 μL) was subjected to cross-linking in solution by irradiation at 350 nm from a distance of 5 cm, at room temperature, for 15 min and 30 min time intervals. A hand-held 100W Blak Ray lamp was used. All cross-linked samples (approx 2μg of protein) were analyzed by SDS-PAGE.

Tryptic analysis and mass spectrometry: Gel bands were cut into small cubes and destained twice with 100 mM NH_4HCO_3 in 50% MeCN (200 μL) for 45 min at 37 ^oC. The gel pieces were dried by Speedvac, and then 10 mM DTT in 100 mM NH_4HCO_3 (100 μL) was added for 15 min at 37 ^oC. The supernatant was removed and 55 mM iodoacetamide in 100 mM NH_4HCO_3 (100 μL) was added for 30 min at 37 ^oC, in the dark. The supernatant was discarded and the gel cubes were washed 3 times with 25 mM NH_4HCO_3 (200 μL) for 15 min at room temp. The gel pieces were dried by speed vac, and then trypsin buffer (12.5 ng/μL trypsin in 50 mM NH_4HCO_3, 60 μL) was added and incubated for 1 hr at 4 ^oC. The supernatant was removed and the gel pieces were covered with 40% (v/v) MeCN in 40 mM NH_4HCO_3 (50 μL) then incubated for 18 hr at 37 ^oC. The supernatant was removed, and saved, then 0.1% TFA (50 μL) was added to the gel pieces for 1 hr at 37 ^oC. This supernatant was combined with the previously saved supernatant and dried by speed vac. The tryptic fragments were resuspended in 50% (v/v) MeCN/0.05% (v/v) TFA (15 μL), ready for mass spectrometry analysis.

MALDI samples were prepared with external standards: insulin (bovine), 25 pmol/μL in 0.1% TFA; insulin oxidized B, 25 pmol/μL in 0.1% (v/v) TFA. Sinapinic acid was used for the matrix at a concentration of 10 mg/mL in 50% (v/v) acetonitrile/0.05% (v/v) TFA. Spotting solutions were made at 4:1:1:2 μL (tryptic protein: insulin ox B: insulin: matrix) of which, 2 μL were spotted on the MALDI plate, in triplicate. MALDI spectra were recorded on an Axima-CFR spectrometer (Shimadzu) operating in reflector mode at 60-90% maximum intensity, optimized for 3500. All spectra were the sum of 500 laser shots per spectra. Reported masses are the monoisotopic masses.

Sequence of protein (GST) used in this study

MDPSSRSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLE[F]P
NLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSK
DFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCL
DAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVSRGVDHH
HHHH

Bracketed "F" mutated to pBpa.

Sequence numbering in the main text refers to wild type GST sequence. The protein used in our studies contains a MDPSSR N-terminal leader peptide and a C-terminal HHHHHH peptide.

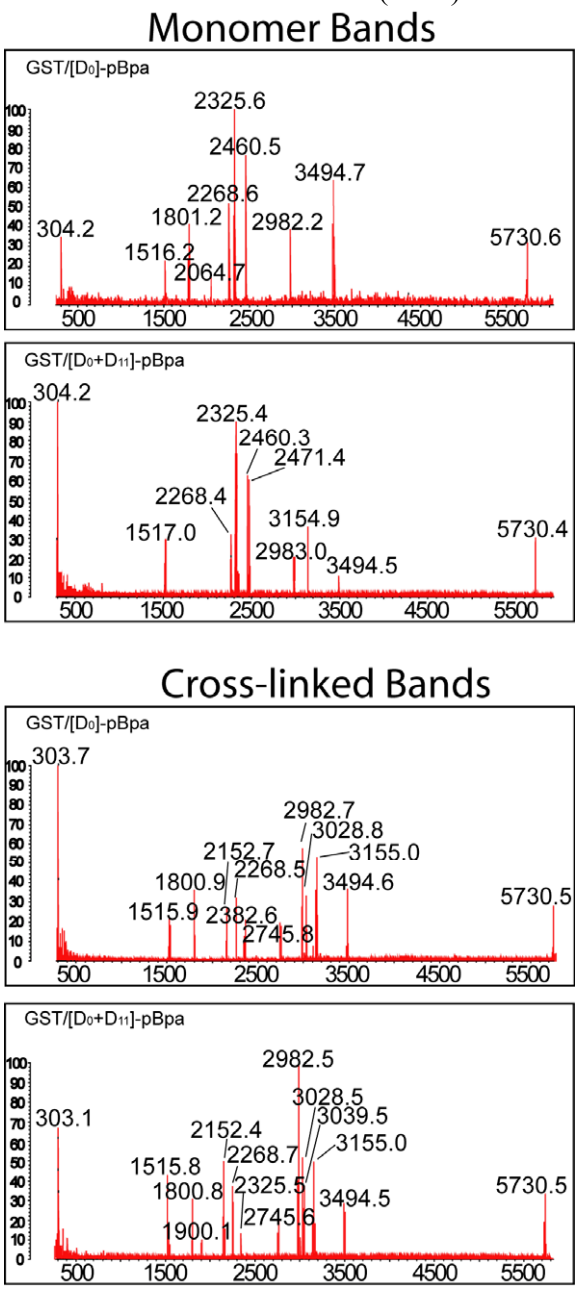
Peptide searching

Peptide fragment peaks from the MALDI-TOF data were analyzed by MASCOT (www.matrixscience.com) under the following search parameters: The database searched was MSDB; taxonomy was open to all entries; peptide mass tolerance was set to ±1.2 Da; carbamidomethyl (C) was selected for fixed modifications; and the number of missed cleavages was set to 3. The mass of the peptide bearing the unnatural amino acid was manually adjusted. In all cases, the top scoring protein was GST.

Monomer Bands	Mascot Score	Sequence Coverage	Matched Peptide Sequences	Observed Mass
[D0]-pBpa	65	33	Unassigned	2982.2
			KFELGLE(pBpa)PNLPYYIDGDVK	2460.5
			YIAWPLQGWQATFGGGDHPPK	2325.6
			LLLEYLEEKYEEHLYER	2268.6
			Unassigned	2064.7
			ERAEISMLEGAVLDIR	1801.2
			AEISMLEGAVLDIR	1516.2
			ER	304.2
[D11]-pBpa	88	38	LLLEYLEEKYEEHLYERDEGDKWR	3155.0
			Unassigned	2981.9
			KFELGLE(pBpa)PNLPYYIDGDVK	2471.4
			YIAWPLQGWQATFGGGDHPPK	2325.5
			LLLEYLEEKYEEHLYER	2269.7
			YEEHLYERDEGDKWRNK	2267.6
			ERAEISMLEGAVLDIR	1801.1
			AEISMLEGAVLDIR	1516.0
			ER	304.2
[D0]-pBpa + [D11]-pBpa	72	38	LLLEYLEEKYEEHLYERDEGDKWR	3154.9
			Unassigned	2983.0
			KFELGLEpBpaPNLPYYIDGDVK	2460.3 /2471.4
			YIAWPLQGWQATFGGGDHPPK	2325.4
			YEEHLYERDEGDKWRNK	2269.7
			LLLEYLEEKYEEHLYER	2268.4
			AEISMLEGAVLDIR	1517.0
			ER	304.2

Cross-linked Bands	Mascot Score	Sequence Coverage	Matched Peptide Sequences	Observed Mass
[D0]-pBpa	60	36%	LLLEYLEEKYEEHLYERDEGDKWR	3155.0
			FELGLE(pBpa)PNLPYYIDGDVK -- MFEDR (cross-linked)	3028.8
			Unassigned	2982.7
			Unassigned	2745.8
			Unassigned	2382.6
			YIAWPLQGWQATFGGGDHPPK	2325.3
			LLLEYLEEKYEEHLYER	2268.5
			DFETLKVDFLSKLPEMLK	2152.7
			ERAEISMLEGAVLDIR	1800.9
			AEISMLEGAVLDIR	1515.9
			ER	303.7
[D11]-pBpa	66	36%	LLLEYLEEKYEEHLYERDEGDKWR	3155.0
			FELGLE(pBpa)PNLPYYIDGDVK -- MFEDR (cross-linked)	3039.5
			Unassigned	2982.5
			Unassigned	2745.8
			YIAWPLQGWQATFGGGDHPPK	2325.2
			LLLEYLEEKYEEHLYER	2268.3
			DFETLKVDFLSKLPEMLK	2152.1
			ERAEISMLEGAVLDIR	1800.8
			AEISMLEGAVLDIR	1516.0
			ER	303.4
[D0]-pBpa + [D11]-pBpa	75	40%	LLLEYLEEKYEEHLYERDEGDKWR	3155.0
			FELGLE(pBpa)PNLPYYIDGDVK -- MFEDR (cross-linked)	3028.5/3039.5
			Unassigned	2982.5
			Unassigned	2745.8
			YIAWPLQGWQATFGGGDHPPK	2325.5
			LLLEYLEEKYEEHLYER	2268.7
			DFETLKVDFLSKLPEMLK	2152.4
			GLVQPTRLLLEYLEEK	1900.1
			ERAEISMLEGAVLDIR	1800.8
			AEISMLEGAVLDIR	1515.8
			ER	303.1

Full Mass Spectra of samples shown in Figure 1.
Internal Standards are insulin (5730) and insulin oxB (3494)



References

Supplementary Material (ESI) for *Molecular BioSystems*

1. S. M. Lamos, C. J. Krusemark, C. J. McGee, M. Scalf, L. M. Smith and P. J. Belshaw, *Angew Chem Int Ed Engl*, 2006, **45**, 4329-4333.